Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 5167–5171, June 1993
Biochemistry

The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors

(protein structure/x-ray crystallography/cytokine structure/colony-stimulating factor 3)

CHRISTOPHER P. HILL*, TIMOTHY D. OSSLUND†, AND DAVID EISENBERG

Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024-1569

Contributed by David Eisenberg, December 8, 1992

ABSTRACT **We have determined the three-dimensional structure of recombinant human granulocyte-colony-stimulating factor by x-ray crystallography. Phases were initially obtained at 3.0-Å resolution by multiple isomorphous replacement and were refined by solvent flattening and by averaging of the electron density of the three molecules in the asymmetric unit. The current *R* factor is 21.5% for all data between 6.0- and 2.2-Å resolution. The structure is predominantly helical, with 104 of the 175 residues forming a four-α-helix bundle. The only other secondary structure is also helical. In the loop between the first two long helices a four-residue 3_{10}-helix is immediately followed by a 6-residue α-helix. Three residues in the short connection between the second and third bundle helices form almost one turn of left-handed helix. The up–up–down–down connectivity with two long crossover connections has been reported previously for five other proteins, which like granulocyte-colony-stimulating factor are all signaling ligands: growth hormone, granulocyte/macrophage-colony-stimulating factor, interferon β, interleukin 2, and interleukin 4. Structural similarity among these growth factors occurs despite the absence of similarity in their amino acid sequences. Conservation of this tertiary structure suggests that these different growth factors might all bind to their respective sequence-related receptors in an equivalent manner.**

Growth and differentiation of various blood cell lines from progenitor stem cells are regulated by a group of proteins known as hematopoietins (1). These proteins include the interleukins (ILs), erythropoietin, macrophage-colony-stimulating factor, granulocyte/macrophage-colony-stimulating factor (GM-CSF), and granulocyte-colony-stimulating factor (G-CSF).

G-CSF is a 19.6-kDa glycoprotein consisting of 174 amino acid residues (2). In G-CSF from human blood, there is one O-linked glycosyl group at Thr^{133} (3), which protects the molecule from aggregation but does not appear to influence receptor binding directly (4). G-CSF, produced mainly by macrophages, induces proliferation of neutrophil colonies and differentiation of precursor cells to neutrophils, and it stimulates the activity of mature neutrophils (5).

G-CSF belongs to a group of growth factors that have been predicted to share a common architecture, despite very low sequence similarity (6, 7). The structures of five of these have been determined—namely, growth hormone (GH) (8, 9), GM-CSF (10, 36), interferon β (IFN-β) (11), IL-2 (12, 13), and IL-4 (14–16); they all have the same four-α-helix bundle motif with up–up–down–down connectivity. Other signaling ligands that are predicted (6, 7) to share this fold include prolactin; erythropoietin; IL-3, -5, -6, and -7; myelomonocytic growth factor (MGF); cholinergic differentiation factor; ciliary neurotrophic factor; and oncostatin M. Only MGF and IL-6 show any significant sequence similarity to G-CSF, with 37% and 32% conservation of sequence identity, respectively (7).

We have determined the crystal structure of recombinant human G-CSF (rhG-CSF),‡ which is expressed in *Escherichia coli*, is not glycosylated and retains the amino-terminal fMet residue; it is, however, biologically active (2). We refer to the fMet as residue number −1.

MATERIALS AND METHODS

The rhG-CSF used in this study was provided by Amgen Biologicals. Crystals were grown (T.D.O., R. Lüthy, D. Cascio, and D.E., unpublished work) in hanging drops at pH 5.8 over a reservoir of 8% (wt/vol) PEG 8000/380 mM $MgSO_4$/220 mM LiCl. The space group is $P2_12_12_1$ (a = 91.2 Å, b = 110.3 Å, c = 49.5 Å). There are three molecules in the asymmetric unit and the Matthews parameter V_m (17) is 2.1 $Å^3$/Da.

Soaking experiments were performed by dissolving the heavy-atom compound in reservoir solution, a small volume of which was added to drops containing the crystals. Data from native and heavy-atom-soaked crystals were collected to 3.0-Å resolution on a San Diego Multiwire Systems area detector. Data were processed using the program FS (18), and two heavy-atom derivatives, thimerosal and praseodymium acetate, were used to initiate determination of the protein structure (details are given in Tables 1 and 2).

The solvent-flattened (21) multiple isomorphous replacement (MIR) map clearly showed the presence of four helices corresponding to each of the three molecules in the asymmetric unit. These three molecules are not related to each other by a proper rotation axis. Noncrystallographic symmetry operators were derived from a partial atomic model, and averaging of the electron density led to a significant improvement in the quality of the map; revealing longer helices, side-chain density, and parts of the connecting loops. This map was interpreted conservatively as 100 alanine residues in the helical regions only. Positional refinement with PROLSQ (22) gave an *R* factor of 38%. Combination of MIR and polyalanine model phases prior to solvent leveling and averaging gave a map that in most regions was of good quality, and into which it was easy to fit the known amino acid sequence (23).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: G-CSF, granulocyte-colony-stimulating factor; rhG-CSF, recombinant human G-CSF; GM-CSF, granulocyte/macrophage-colony-stimulating factor; IL, interleukin; IFN, interferon; GH, growth hormone; hGH, human GH; pGH, porcine GH; MIR, multiple isomorphous replacement.

*Present address: Department of Biochemistry, University of Utah, Salt Lake City, UT 84132.
†Present address: Amgen, Molecular Structure Lab, Building 2, Thousand Oaks, CA 91360.

‡The atomic coordinates and structure factors have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 1RHG, R1RHGSF).

Table 1. Data used to solve the structure of rhG-CSF

	Native	Thimerosal	$Pr(OAc)_3$
Soaking		0.1 mM, 48 hr	0.1 mM, 48 hr
No. of sites		3	4
R_{iso}*		0.21	0.24
Resolution	2.9 Å	3.0 Å	3.0 Å
No. of observations	50,868	46,840	43,495
No. of rejects	1,281	1,565	5,970
No. unique	10,735	10,031	8,815
% completeness	93.3	96.1	84.4
% of $\|F\| > 2\sigma F$	95.5	97.0	96.7
R_{sym} (no rejects)†	0.056	0.057	0.071

Unless otherwise stated, crystallographic programs used were from the CCP4 suite (19). $Pr(OAc)_3$, praseodymium acetate.
*$R_{iso} = \Sigma||F_{derivative}| - |F_{native}||/\Sigma|F_{native}|$.
†$R_{sym} = \Sigma|I_i - I_{av}|/\Sigma I_{av}$.

Most of the structure, especially the helical regions, was clearly defined at this stage. Other sections were less precisely determined; in particular, residues −1 to 9, 65 to 70, 127 to 136, and 173 to 174 did not have convincing density. In general the segments that could not be located in MIR/averaged maps are close to neighboring molecules in the crystal, suggesting that these regions of poor electron density may be due to averaging out features that differ among the three molecules. Consequently, a mask was manually defined to include only regions that were known to obey the noncrystallographic symmetry (i.e., were well defined in the averaged map). Another mask was defined for regions that were confidently expected to be solvent (i.e., well removed from the missing residues). An iterative procedure of map calculation, modification, and back-transformation was employed, in which regions outside the masks were not modified explicitly. Several different strategies were tried, all with similar results (Fig. 1). This gave a noticeable improvement in the map quality; however, it was still not possible to locate the missing residues, which remain undefined even after refinement.

The structure has been refined with XPLOR (25) against 2.2-Å data collected on an *R*-axis imaging plate detector. The current *R* factor is 21.5% and the rms deviation from ideality of covalent bond lengths is 0.017 Å; further details are given in Tables 3 and 4. The unusually large average atomic *B* factor for protein atoms, 44 $Å^2$, is in very good agreement with the value estimated from a Wilson plot (26), 45 $Å^2$. The accuracy of our model is supported by the free *R* factor (27), which is 34.4%, and by three-dimensional profile assessment (28).

FIG. 1. MIR/averaged density for helix D of rhG-CSF. This map was computed from F_o coefficients with MIR phases that were refined by 10 cycles of averaging, Fourier inversion, and map calculation with combined MIR/averaged phases. Solvent leveling was not applied in this case. Map averaging was performed using Bricogne's programs (24). The helical main chain is well defined and in general the side chains have reasonable density. Side chains with weak or absent density are usually fully exposed to solvent—for example, Glu^{162}, Tyr^{165}, and Arg^{169}.

RESULTS

rhG-CSF is an antiparallel four-α-helix bundle with a left-handed twist, and with overall dimensions of 45 Å × 26 Å × 26 Å (Fig. 2). The four helices within the bundle are referred to as helices A–D; their connecting loops are known as the AB, BC, and CD loops. The AB and CD loops are long overhand connections; only the BC loop is of the more usual short hairpin type (31).

The rhG-CSF bundle is regular with helix crossing angles that range between −167° and −159°. The average crossing angle (−162.5°) is very close to that expected (−161°) for an ideal left-handed antiparallel four-α-helix bundle (32). Helices A, B, and C are straight, whereas helix D bends towards the shorter helix B. The change in axial direction between the ends of helix D is 35°, with the greatest changes centered on Gly^{149} and Ser^{159}. The longest straight portion of helix D (residues 159–173) makes the most extensive interactions with the A, B, and C helices, and this is the section of helix D used above to define the crossing angles within the bundle.

In addition to the four major helices that comprise the bundle, there is a shorter helical section within the AB loop.

Table 2. Heavy-atom phasing statistics

Resolution	9.91	6.56	5.22	4.46	3.96	3.60	3.32	3.09	Total
Thimerosal									
$f_{h(rms)}/E_{(rms)}$(centrics)	0.93	1.04	0.74	0.62	0.64	0.60	0.67	1.00	0.77
$f_{h(rms)}/E_{(rms)}$(acentrics)	1.13	1.35	1.03	0.83	0.84	0.85	0.94	1.08	0.98
R_{Cullis}	0.59	0.57	0.75	0.71	0.68	0.70	0.72	0.57	0.66
$Pr(OAc)_3$									
$f_{h(rms)}/E_{(rms)}$(centrics)	1.13	0.94	0.75	0.43	0.44	0.61	0.53	0.60	0.66
$f_{h(rms)}/E_{(rms)}$(acentrics)	1.31	1.38	0.92	0.64	0.70	0.79	0.77	0.80	0.84
R_{Cullis}	0.52	0.59	0.67	0.72	0.77	0.69	0.75	0.72	0.67
Mean figure of merit = 0.59									

Heavy-atom parameters were refined by the method of correlating origin-removed Patterson functions (20). The phasing power of the praseodymium acetate [$Pr(OAc)_3$] derivative, 0.84, would normally indicate a useless derivative. In this case, however, the density maps were clearly improved when anomalous scattering from $Pr(OAc)_3$ was included in the phase calculation. We assume that this results from the large anomalous signal ($f'' = 10.5e$) of praseodymium. The heavy-atom binding sites are chemically reasonable. Each of the three thimerosal mercury atoms is bound to the single free thiol, Cys^{17}, in each of the three molecules in the asymmetric unit. In contrast, praseodymium sites do not obey the noncrystallographic symmetry. They are located near clusters of at least three carboxylate side chains, which in every case come from at least two different rhG-CSF molecules. f_h = heavy-atom structure factor; *E* = residual lack of closure; $R_{Cullis} = \Sigma||F_{PH} - F_P| - |f_{h(calc)}||/\Sigma|F_{PH} - F_P|$.

Table 3. Refinement of rhG-CSF

Resolution, Å	*R* factor	Free *R* factor
6.00–3.99	16.3	34.2
3.99–3.34	17.7	33.1
3.34–2.97	22.4	32.7
2.97–2.73	26.4	35.6
2.73–2.55	28.2	36.5
2.55–2.41	30.0	34.8
2.41–2.29	31.3	38.5
2.29–2.20	32.7	37.2
Overall	21.5	34.4

The protein chain makes a sharp turn out of helix A and 5 residues later goes into a 4-residue 3_{10}-helix. At Leu[47] there is a shift in the chain direction and the 3_{10}-helix leads immediately into a 6-residue α-helix. The 45° angle between these short helices wraps them around the N-terminal end of helix D. They are relatively exposed and protrude from the main body of the structure. Residues from this region overlap with epitopes recognized by neutralizing monoclonal antibodies (33).

All of the non-glycine residues fall into the allowed region of ϕ,ψ space, except for Lys[40], Glu[93], and Ile[95]. Lys[40] lies at the C terminus of helix A, a context that is known to favor the α_L conformation (34). Glu[93], Gly[94], and Ile[95], located in the short BC loop, all have positive ϕ angles and together form almost one turn reminiscent of a left-handed helix. The conformation of these residues is clear in the MIR/averaged map.

The two disulfide bonds in rhG-CSF, Cys[36]–Cys[42] and Cys[64]–Cys[74], are both required for activity (35). They are located at opposite ends of the long AB loop, where they form short loops to the C-terminal end of helix A and the N-terminal end of helix B (Figs. 2 and 3). The Cys[36]–Cys[42] disulfide forms the major part of a neutralizing antibody epitope (33). Circular dichroism measurements show that in the absence of Cys[64]–Cys[74] only about half of the native structure α-helix is formed (35).

DISCUSSION

rhG-CSF belongs to a distinct structural class of growth factors. Comparison with GM-CSF (10, 36), GH (8, 9), IFN-β (11), IL-2 (12), and IL-4 (14–16) reveals a common motif of a four-α-helix bundle with two long crossover connections. This similarity is evident from connectivity diagrams (Fig. 3), and it exists despite little sequence similarity. Differences include the number and position of disulfide bonds. To our knowledge no other proteins are yet known to share this architecture.

Table 4. The final model

Residues included	
Molecule A	9–61, 71–126, 137–172
Molecule B	209–264, 271–322, 337–372
Molecule C	409–461, 471–526, 537–572
No. of water molecules	40
rms deviations	
Bonds	0.017 Å
Angles	3.354°
Dihedrals	20.805°
Impropers	1.437°
B (main-chain bonds)	1.63 Å

The residue numbers of molecules B and C have been incremented by 200 and 400, respectively. Nine residues that lack clear side-chain density have been truncated to Ala.

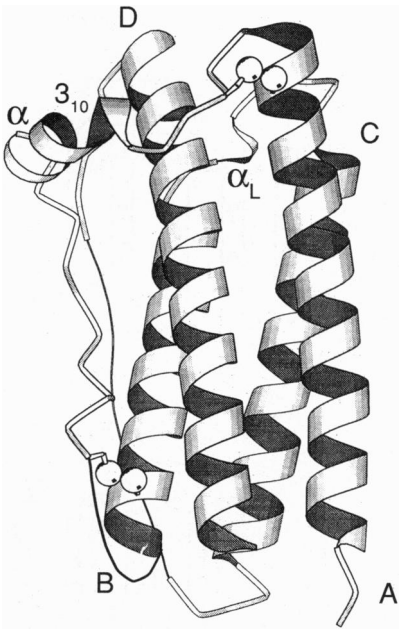

FIG. 2. Ribbon diagram of rhG-CSF. This figure was prepared using the program MOLSCRIPT (29). Secondary structure was defined using DSSP (30). The main bundle helices A (residues 11–39), B (71–91), C (100–123), and D (143–172) are labeled near their N termini. The short 3_{10} (44–47) and α (48–53) helices are also indicated, as are residues 93–95, which form almost one turn reminiscent of a left-handed helix. Residues −1 to 8 and 173 to 174 are not visible in Fourier maps and are not included in this figure. The approximate positions of the other missing residues, 65–70 and 127–136, have been drawn with thin lines to indicate connectivity.

These growth factors differ in the local conformations of their loops. For example the BC loops of IL-4 and hGH contain short α-helices, while pGH has an ω loop (37), and rhG-CSF has almost one turn of a left-handed helix in this position. The AB and CD loops of GM-CSF and IL-4 form two-stranded antiparallel β-sheets that run approximately parallel to the bundle axis; the CD loop of IFN-β contains a 17-residue α-helix that is aligned with the bundle axis and packs against the helices equivalent to B and D of rhG-CSF. Only hGH, pGH, and IL-2 appear to have helices that are even approximately close to the short 3_{10}- and α-helices found in the AB loop of rhG-CSF.

There are also differences in the bundle geometries. The bundle helices of rhG-CSF, pGH, and hGH have almost twice as many residues as in GM-CSF. The relative lengths of bundle helices vary (Fig. 3): for example, in rhG-CSF, pGH, and hGH the A and D helices are longer than the B and C helices, whereas for IL-4 the reverse is true. The helix crossing angles also differ. rhG-CSF, which has one of the longest bundles, has crossing angles very close to those expected for packing of $i + 4$ ridges into $i + 3$ grooves (−160°). At the other extreme, GM-CSF, which has the smallest bundle, has wide crossing angles. IL-4, whose bundle is intermediate in size, has crossing angles that are intermediate between those of rhG-CSF and GM-CSF.

Despite their differences, these growth factors, which all function by binding to cell surface receptors, clearly share the same basic architecture. The extracellular portion of their receptors includes a conserved "cytokine-binding" domain of ≈210 residues (38–42). This suggests that these signaling ligands might all bind to their respective receptors with equivalent geometries. Thus far, hGH is the only one of these ligands whose receptor-bound structure has been reported (9). A single molecule of hGH facilitates receptor dimerization by simultaneously binding to two different receptor

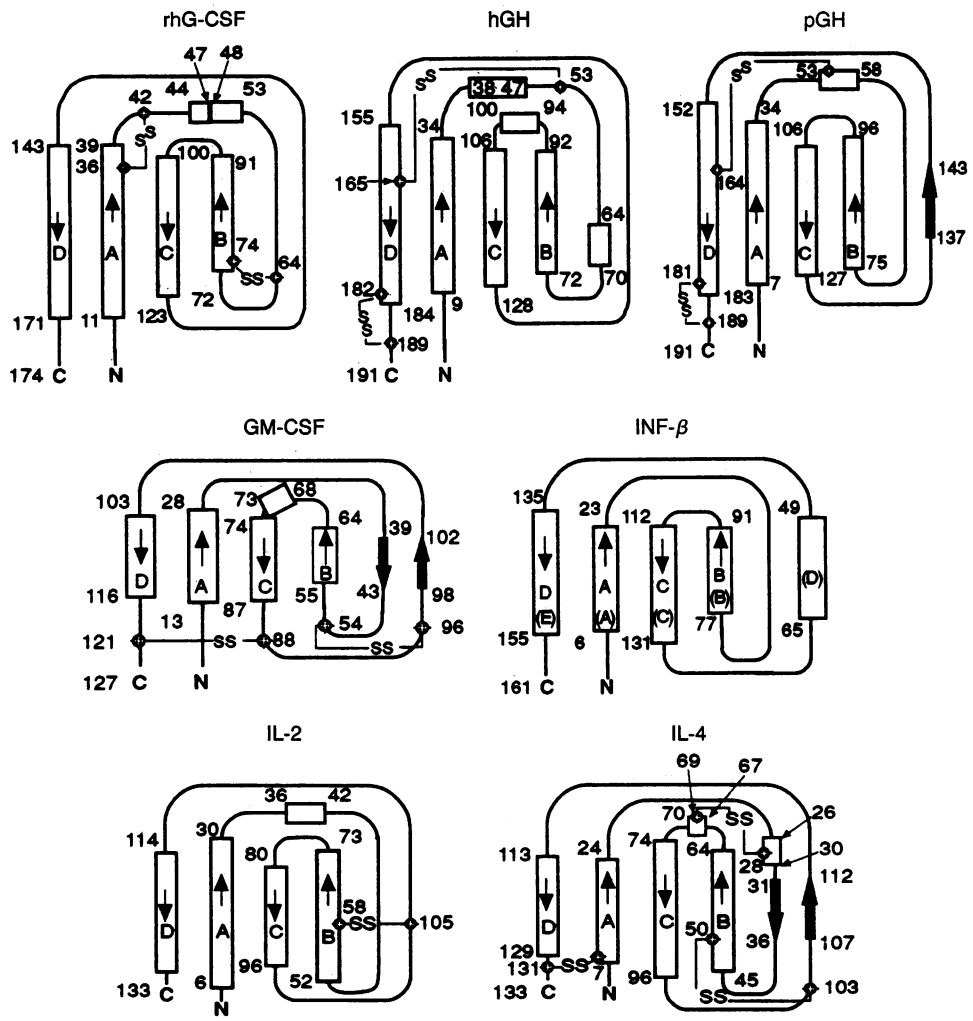

FIG. 3. Connectivity diagrams of rhG-CSF, human GH (hGH; ref. 9), porcine GH (pGH; ref. 8), GM-CSF (10, 36), IFN-β (11), IL-2 (12), and IL-4 (14, 15). These diagrams are based on inspection of cited references. The lengths of secondary structural elements are drawn in proportion to the number of residues. There are some minor differences in the secondary structure assignments reported by the various groups that have independently determined the structures of GM-CSF and of IL-4. A, B, C, and D helices are labeled according to the scheme used in this paper for rhG-CSF. For IFN-β the original labeling of helices is indicated in parentheses.

molecules. The first hGH receptor-binding surface is primarily composed of residues from helix A, the AB loop, and helix D; the residues of the second binding surface are from helix A, the BC loop, and helix C. It remains to be seen whether the structural similarity shown within this class of growth factors extends to their mode of receptor recognition.

Note Added in Proof. The recently reported crystal structure reveals that macrophage-colony-stimulating factor (M-CSF) (43) belongs to the same structural family as G-CSF, although M-CSF is unique in consisting of disulfide-linked dimers. In this regard we note that IFN-γ is a dimer in which interpenetrating helices form two domains, each of which is reminiscent of IFN-β (44).

We thank the National Institutes of Health for support and the San Diego Super Computer for computer time.

1. Nicola, N. A. (1989) *Annu. Rev. Biochem.* **58**, 45–77.
2. Souza, L. M., Boone, T. C., Gabriloe, J., Lai, P. H., Zsebo, K. M., Murdock, D. C., Chazin, V. R., Bruszewski, J., Lu, H., Chen, K. K., Barendt, J., Platzer, E., Moore, M. A. S., Mertelsmann, R. & Welte, K. (1986) *Science* **232**, 61–65.
3. Kubota, N., Orita, T., Hattori, K., Oh-eda, M., Ochi, N. & Yamazaki, T. (1990) *J. Biochem. (Tokyo)* **107**, 486–492.
4. Oh-eda, M., Hasegawa, M., Hattori, K., Kuboniwa, H., Kojima, T., Orita, T., Tomonou, K., Yamazaki, T. & Ochi, N. (1990) *J. Biol. Chem.* **265**, 11432–11435.
5. Nagata, S. (1990) in *Handbook of Experimental Pharmacology: Vol 95/I: Peptide Growth Factors and Their Receptors*, eds. Sporn, M. B. & Roberts, A. B. (Springer-Verlag, New York), pp. 699–722.
6. Bazan, J. F. (1990) *Immunol. Today* **11**, 350–354.
7. Bazan, J. F. (1991) *Neuron* **7**, 197–208.
8. Abdel-Meguid, S. S., Shieh, H.-S., Smith, W. W., Dayringer, H. E., Violand, B. N. & Bentle, L. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6434–6437.
9. Vos, A. M., Ultsch, M. & Kossiakoff, A. A. (1992) *Science* **255**, 306–312.
10. Diederichs, K., Boone, T. & Karplus, P. A. (1991) *Science* **254**, 1779–1782.
11. Senda, T., Shimazu, T., Matsuda, S., Kawano, G., Shimizu, H., Nakamura, K. T. & Mitsui, Y. (1992) *EMBO J.* **11**, 3193–3201.
12. McKay, D. B. (1992) *Science* **257**, 412–413.
13. Bazan, J. F. (1992) *Science* **257**, 410–412.
14. Powers, R., Garrett, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M. & Clore, G. M. (1992) *Science* **256**, 1673–1677.
15. Smith, L. J., Redfield, C., Boyd, J., Lawrence, G. M. P., Edwards, R. G., Smith, R. A. G. & Dobson, C. M. (1992) *J. Mol. Biol.* **224**, 899–904.
16. Wlodawer, A., Pavlovsky, A. & Gustchina, A. (1992) *FEBS Lett.* **309**, 59–64.
17. Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491–497.

18. Weissman, L. J. (1979) Ph.D. dissertation (Univ. of California, Los Angeles).
19. The SERC Collaborative Computing Project No. 4 (1979) *A Suite of Programs for Protein Crystallography* (Daresbury Lab., Warrington, U.K.).
20. Terwilliger, T. C. & Eisenberg, D. (1983) *Acta Crystallogr. Sect. A* **39,** 813–817.
21. Wang, B.-C. (1985) *Methods Enzymol.* **115,** 90–112.
22. Hendrickson, W. A. (1985) *Methods Enzymol.* **115,** 252–270.
23. Jones, T. A. (1985) *Methods Enzymol.* **115,** 157–171.
24. Bricogne, G. (1976) *Acta Crystallogr. Sect. A* **A32,** 832–847.
25. Brünger, A. T. (1992) X-PLOR, A System for Crystallography and NMR (Yale Univ., New Haven, CT), Version 3.0.
26. Wilson, A. J. C. (1949) *Acta Crystallogr.* **2,** 318–321.
27. Brünger, A. T. (1992) *Nature (London)* **355,** 472–475.
28. Lüthy, R., Bowie, J. U. & Eisenberg, D. (1992) *Nature (London)* **356,** 83–85.
29. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
30. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22,** 2577–2637.
31. Presnell, S. R. & Cohen, F. E. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6592–6596.
32. Chothia, C., Levitt, M. & Richardson, D. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4130–4134.
33. Layton, J. E., Morstyn, G., Fabri, L. J., Reid, G. E., Burgess, A. W., Simpson, R. J. & Nice, E. C. (1991) *J. Biol. Chem.* **266,** 23815–23823.
34. Richardson, J. S. & Richardson, D. C. (1988) *Science* **240,** 1648–1652.
35. Lu, H. S., Clogston, C. L., Narhi, L. O., Merewether, L. A., Pearl, W. R. & Boone, T. C. (1992) *J. Biol. Chem.* **267,** 8770–8777.
36. Walter, M. R., Cook, W. J., Ealick, S. E., Nagabhushan, T. L., Trotta, P. P. & Bugg, C. E. (1992) *J. Mol. Biol.* **224,** 1075–1085.
37. Leszczynski, J. F. & Rose, G. D. (1986) *Science* **234,** 849–855.
38. Cosman, D., Lyman, S. D., Idzerda, R. L., Beckmann, M. P., Park, L. S., Goodwin, R. G. & March, C. J. (1990) *Trends Biochem. Sci.* **15,** 265–270.
39. Arai, K.-I., Lee, F., Miyajima, A., Miyatake, S., Arai, N. & Yokota, T. (1990) *Annu. Rev. Biochem.* **59,** 783–836.
40. Bazan, J. F. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6934–6938.
41. Davis, S., Aldrich, T. H., Valenzuela, D. M., Wong, V., Furth, M. E., Squinto, S. P. & Yancopoulos, G. D. (1991) *Science* **253,** 59–63.
42. Gearing, D. P., Thut, C. J., Vandenbos, T., Gimpel, S. D., Delaney, P. B., King, J., Price, V., Cosman, D. & Beckmann, M. P. (1991) *EMBO J.* **10,** 2839–2848.
43. Pandit, J., Bohm, A., Jancarik, J., Halenbeck, R., Koths, K. & Kim, S.-H. (1992) *Science* **258,** 1358–1362.
44. Ealick, S. E., Cook, W. J., Vijay-Kumar, S., Carson, M., Nagabhushan, T. L., Trotta, P. P. & Bugg, C. E. (1991) *Science* **252,** 698–702.